Available online www.jsaer.com

***Journal of Scientific and Engineering Research*, 2016, 3(4):295-301**

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

EpiOne: A Software Tool for Identification of Potential Cytosine DNA Methylation Marks in Promoters and Gene Bodies

Mehmet Karaca, Ayse Gul Ince

Department of Field Crops, Faculty of Agriculture, Akdeniz University, Turkey

Abstract DNA cytosine methylation is one of the most important epigenetic marks. To date a great number of software tools have been developed to detect, visualize or quantify DNA cytosine methylation data obtained from chemical treatment or enzyme converted genomic sequences and microarray data. However, there exist no software tools specifically identifying potential DNA cytosine methylation marks (CpG, CpHpG, and CpHpH) in promoters and gene bodies. Here we report several algorithms embedded in a novel software tool, EpiOne, to identify and locate potential cytosine methylation marks in annotated genomic DNA sequences. Algorithms also could be used to extract DNA sequences into promoter, exon, intron and UTR sequences. Algorithms of the EpiOne could be used to investigate relationships between some gene properties such as gene, exon, intron lengths and potential cytosine methylation marks and gene expression. Furthermore algorithms identify candidate genomic regions for primer binding sites or target sequences to be used in bisulfite sequencing studies, which is the gold standard in epigenetic research.

Keywords bisulfite sequencing, data mining, exon, intron, promoter

Introduction

First coined by Conrad Hal Waddington in the middle of 20th century and referred to "changes in the state of cell differentiation during development, and the way in which cell fates become restricted as development proceeds" [1], epigenetics literally means "on top of" or "in addition" to genetics. It is the study of *meiotically and mitotically* heritable variations in gene expression differences that are not controlled by changes in DNA sequence [2]. Gene expression variation caused by epigenetics includes chromatin structure and packaging, gene expression, genomic imprinting and recombination [3-6]. Epigenetic instability in human causes cancers and some other diseases including hepatic cancer, gastric cancer, colorectal cancer, prostate cancer [7-11]. Epigenetic instability in plants causes stunting, abnormal leaf and flower, decreased apical dominance and reduced fertility [12-15].

Epigenetic variation occurs in histones (methylation, phosphorylation, acetylation etc,), DNA (5-methylcytosine, 5-hydroxymethylcytosine) and RNA (N^6-methyladenosine and N^6-hydroxymethyladenosine [16,17]. Among the epigenetic modifications, DNA methylation was the first discovered epigenetic mark. DNA cytosine methylation involves the addition of a methyl group to the carbon-5 position of cytosine residues in the CpG, CpHpG or CpHpH contexts (where H refers As, Ts, or Cs) of pluripotent mammalian cells and plant cells. DNA methyltransferases (DNMTS) are enzymes responsible for DNA cytosine methylation. Several DNMTS are currently known, each with distinct roles. Some DNMTS involve in the establishment of DNA methylation patterns on unmethylated DNA acting as *de novo* methyltransferases. Some other DNMTS maintain DNA methylation patterns, acting preferentially on hemi-methylated DNA [18].

It is well know that the addition of a methyl group to cytosine induces or reduces gene expressions in animal and plant genomes. For instance, methyl addition can prevent the binding of factors to the DNA that promotes transcriptional activity. Similarly methyl addition can induce gene expression by preventing the binding of insulators to DNA allowing induced gene activity. In short, methylation may causes *cis*, *trans*, and chromatin level changes and results in differential expression of a number of genes [16,19].

The technology to identify and locate DNA cytosine methylation has a steady progress in the last 3 decades. Methods determining DNA cytosine methylation can be divided into 3 main categories; global, genome-wide

and gene-specific DNA cytosine methylation. Global methylation methods determine total DNA cytosine methylation in a genome by utilization of some restriction enzymes, single-nucleotide extension with polymerase chain reaction (PCR), mobility and melting property assays. High-performance liquid chromatography, high-performance capillary electrophoresis, methylation-sensitive single-strand conformation analysis, high resolution melting analysis, methylation-sensitive single-nucleotide primer extension, mass spectrometry and methylation-specific digital karyotyping provide global methylation analysis [15,16,19-22].

Genome-wide methylation methods assess locus-specific DNA cytosine methylation on a genome-wide scale. Methods include restriction landmark genomic scanning, ligation-mediated polymerase chain reaction, methylated CpG island recovery assay, differential methylation hybridization, amplification of inter-methylated sites, HpaII tiny fragments enrichment by ligation-mediated PCR, reduced representation bisulfite sequencing, single-molecule, real time sequencing, methylated DNA immunoprecipitation (MeDIP), microarray chip hybridization (MeDIP-Chip) and methylated DNA immunoprecipitation followed by sequencing (MeDIP-Seq) [16,19,23, 24].

Gene-specific (locus-specific) methods determine DNA methylation patterns of specific genes or gene regions such as gen bodies and promoters. Methods include PCR amplification of DNA material digested with restriction enzymes differentially sensitive to CpG DNA methylation, Southern blot detection of genomic regions from DNA material digested with restriction enzymes, methylation-specific PCR (MSP), methyl light or quantitative analysis of methylated alleles, melting curve analysis, combined bisulfite restriction analysis, digital bisulfite sequencing, bisulfite sequencing and pyrosequencing [16,19].

To date a great number of software tools have been developed to detect, visualize or quantify DNA cytosine methylation data obtained from above mentioned DNA methylation methods [16,19,25,26]. There is limited number of software specifically identifying potential cytosine methylation marks (CpG, CpHpG or CpHpH) in regulator regions such as enhancers, promoters, insulators, silencers and locus control regions, and gene body regions such as upstream and downstream regions (untranslated terminal regions UTRs), transcription start site (TSS), coding exons or introns, transcription termination sites (TTSs).

The rate and pattern of cytosine DNA methylation differ strongly among organisms, tissues and development stages within an organism. Identification of DNA cytosine methylation density and distribution in gene bodies and regulatory sequences will provide us further information about the epigenetic regulation in organisms. Here we report novel algorithms and a software tool EpiOne to identify and locate potential cytosine methylation marks (CpG, CpHpG or CpHpH) in genomic DNA sequences. Algorithms will be useful to extract promoter, exon, intron and UTR sequences useful in epigenetic studies.

Materials and Methods

In order to explain various terminologies we used in the software, we use an example sequence, given by S= atctcccgtaccccgagctt. Let the length of the sequence S be denoted by L_S. L is DNA sequence formatted in a GenBank format. L consists of an annotation section and a sequence section. The annotation section start site is marked by a line “locus” and the sequence start section is marked by a line "origin" and the end of the section is marked by a line "//" [27]. S(n) is used to represent the n-th base of the sequence S. For example, S(0) = ‘a’ and S(1) = ‘t’ in the above given DNA sequence. Algorithms recognizing these marks have been published in [28].

Algorithms

In order to increase the efficiency of algorithms, EpiOne software tool utilizes two types of memories, the main memory, which is fast but costly and scarce, and the secondary memory, which is slow but cheap and abundant. Algorithms adjusted the two types of memories so that the tool EpiOne has an optimal balance between its speed and its main memory requirements.

Two main algorithms are introduced under the processing options for identification of CpG, CpHpG and CpHpH (where H=A, T or C) methylation marks in promoters, introns, exons, 3'-UTRs and 5'-UTRs. In the first type algorithm, EpiOne locates potantiel methylation marks CpG, CpHpG and CpHpH individually or altogether. In this type of serach, potantial metylation marks CpHpG and CpHpH present in the vicinity of dinucleotide CpG are ignored and when H is G CpHpG and CpHpH are reported as CpG. For instance, in a DNA sequence of ATTCCCGTACCCCGAGCTT, CGT string is reported as CpG, not as CpHpH. The second type of serach can be made by selecting the independent search option. In this serach type, each of the potential cytosine methylation mark CpG, CpHpG and CpHpH is located individually.

Algorithms under the promoters & gene body options enable the user to specifically study or collect promoter or any gene bodies indivudually. Algorithms in processing options determine the the type of analysis required by the user. For instance when extract gene elements option is selected, the software will write individual fasta-formated files, each of which consists of promoter, intron, exon, 3'-UTR or 5'-UTR sequences. This algorithm enables users to extract and save promoters, introns, exons 5'-UTR and 3'-UTR DNA sequences into different files to be used in other studies. When extract gene elements option is not selected, the software will not

generate fasta-formated files but will produce reports for methylation status of promoter, intron, exon, 3'-UTR or 5'-UTR sequences. In the processing options there is progress only sequences with gene elements option. When this option is selected the EpiOne software will analyze all the genomic DNA sequences without mining promoters or gene bodies. With this option, user can study genomic DNA sequences that are not annotated. Algorithms recognizing promoters, introns, exons and UTRs have been published in [29].

Programing and availibility

Programing language of EpiOne is C and is written in standard C++. EpiOne is freely available upon request from mkaraca@akdeniz.edu.tr.

Results

EpiOne does not require network connection to function. EpiOne and EpiHelp are compressed in a single zip file and it is ready to be used upon unzipping the file. EpiHelp document contains most of the information given in this manuscript along with some other useful information. It is a standalone software and it can be successfully run in today's any personal computer. EpiOne has main interface (Figure 1) and one option interface (Figure 2).

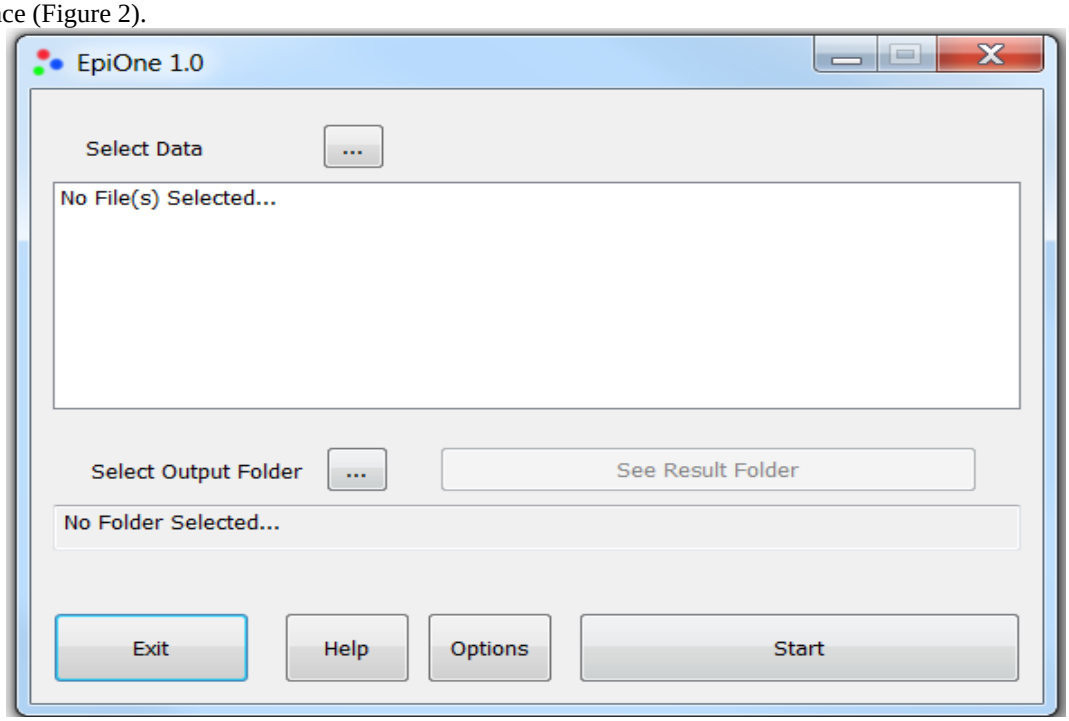


Figure 1: Analysis windows of EpiOne software

In EpiOne option interface there are 3 main options, processing, islet and promoter & gene body options (Figure 2). Processing options can be used to determine potential DNA cytosine methylation distribution and frequency on promoters and gene bodies or extraction of promoters and gene bodies in fasta-formated files to be used in other studies.

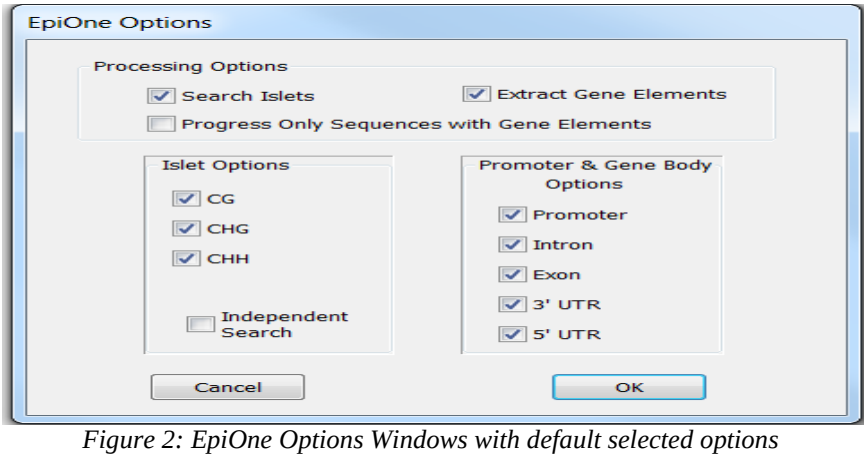


Figure 2: EpiOne Options Windows with default selected options

EpiOne composes of two algorithms for identification of potential DNA cytosine methylation marks. The first one is a search with the selected options through genomic sequences to find potential cytosine methylation marks CpG, CpHpG and CpHpH in genomic DNA sequences. Cytosines in CpG, CpHpG and CpHpH are substrates for DNA methyltransferases [16,18,19].

Under the islet options we introduced another algorithm to detect any cytosine in the promoters and gene bodies of genomic DNA sequences. Conventional algorithms detect cytosine methylation in CpG marks because it is a widespread modification in plant and animal genomes. Many studies reported that CpHpG and CpHpH cytosine marks are also methylated in certain tissues in animal and many plant cells [9,11,15,16]. Therefore EpiOne could be used in any organism's genomic DNA sequences for identification of potential methylation marks.

EpiOne software has popup windows appearing during the analysis informing the progressed sample file names and DNA sequence numbers. Upon completion of the analysis, another popup window with the message "see result folder" appears. Upon completion of the analysis EpiOne software creates a folder in the output folder called as per analysis date. For example a folder named 20150120140850 means that analysis was performed in 2015, 20th January, 14:08:50 o'clock. Depending on the option implemented and the data used, the number of file created by EpiOne software differs. For instance when the default options are used there will be nine files, 5 of which are fasta-formated files created by the software. Generated fasta-formated files include promoter, intron, exon, 5'-UTR and 3'-UTR sequences. In these files, the first line contains information about the organism, GenBank identification number, item no (for example intron or exon number), length and the start and stop position number of the sequence [27].

In the results folder along with the fasta-formated files, there are also summary, islets, error and implemented option report files (Figure 3). Summary file in the result folder can be viewed with excel or other spreadsheet software tools. This file basically summarizes the number of intron, exon, promoter, 5'-UTR sequences and 3'-UTR identified in genomic DNA sequences. This file enables users to compare different genes from the same or different organisms in terms of their intron and exon numbers and the characterized UTR regions.

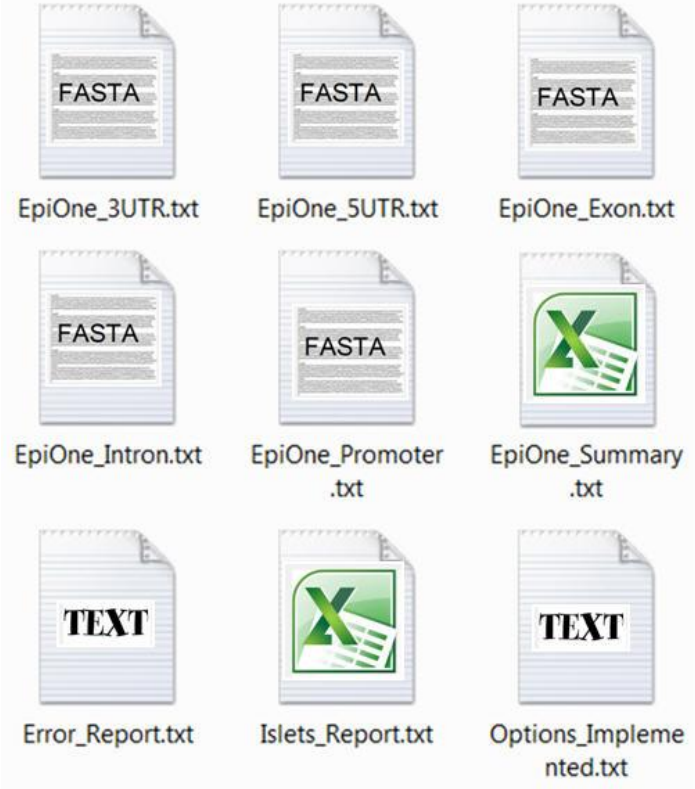
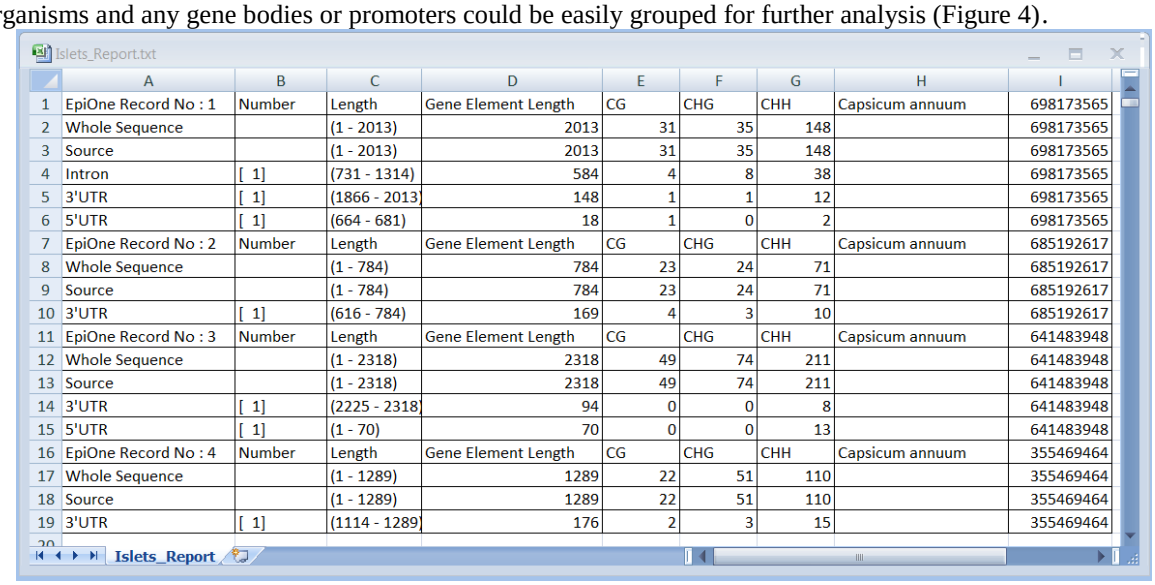


Figure 3: Result folder of EpiOne software

Information about potential methylation information for CpG, CpHpG and CpHpH marks is given in islet report files which can be viewed with excel or other spreadsheet software. Islet report files contain information in nine colons (Figure 4). The first colon comprises information for gene bodies and promoters along with the whole sequence and source gathered from GenBank sequences. The second colon provides information on the number of promoters, 5'-UTRs, exons, introns and 3'-UTRs identified from the genomic DNA sequences. The third colon shows the start and stop of the nucleotide sequence positions while the fourth colon shows the lengths of the gene elements as whole sequence, source, promoter, 5'-UTR, exon, intron, 3'-UTR. The fifth, sixth and

seventh colons show the number of CpG, CpHpG and CpHpH marks identified in the genomic sequences. The eighth colon shows the name of organisms from which the sequences originate. The last colon shows the GI (GenInfo Identifier) for whole sequence, source, promoter, 5'-UTR, exon, intron and 3'-UTR regions. Since the islet report file can be viewed by excel or other spreadsheet software, potential methylation marks for any organisms and any gene bodies or promoters could be easily grouped for further analysis (Figure 4).



Islets_Report.txt

	A	B	C	D	E	F	G	H	I
1	EpiOne Record No : 1	Number	Length	Gene Element Length	CG	CHG	CHH	Capsicum annuum	698173565
2	Whole Sequence		(1 - 2013)	2013	31	35	148		698173565
3	Source		(1 - 2013)	2013	31	35	148		698173565
4	Intron	[1]	(731 - 1314)	584	4	8	38		698173565
5	3'UTR	[1]	(1866 - 2013)	148	1	1	12		698173565
6	5'UTR	[1]	(664 - 681)	18	1	0	2		698173565
7	EpiOne Record No : 2	Number	Length	Gene Element Length	CG	CHG	CHH	Capsicum annuum	685192617
8	Whole Sequence		(1 - 784)	784	23	24	71		685192617
9	Source		(1 - 784)	784	23	24	71		685192617
10	3'UTR	[1]	(616 - 784)	169	4	3	10		685192617
11	EpiOne Record No : 3	Number	Length	Gene Element Length	CG	CHG	CHH	Capsicum annuum	641483948
12	Whole Sequence		(1 - 2318)	2318	49	74	211		641483948
13	Source		(1 - 2318)	2318	49	74	211		641483948
14	3'UTR	[1]	(2225 - 2318)	94	0	0	8		641483948
15	5'UTR	[1]	(1 - 70)	70	0	0	13		641483948
16	EpiOne Record No : 4	Number	Length	Gene Element Length	CG	CHG	CHH	Capsicum annuum	355469464
17	Whole Sequence		(1 - 1289)	1289	22	51	110		355469464
18	Source		(1 - 1289)	1289	22	51	110		355469464
19	3'UTR	[1]	(1114 - 1289)	176	2	3	15		355469464

Islets_Report

Figure 4. Islets report output file of EpiOne software

As for all the software tools, EpiOne has some limitations. Although there is no limit for numbers of DNA sequences in input files, the lenght of a sequence is 1000000 nucleotides. Fortunatelly the lenghts of many know genes are quite below of this limit [30]. In order to use computer resources more efficiently we introduced several limitations on the number of intron, exon, UTR and regulator regions. For instance a gene can contain upto 10000 introns or exons. The upper limit for UTRs or regulatos per gene is 1000. EpiOne can identify upto 40000 CpG, CpHpG or CpHpH, that means a sequence may contain upto 120000 cytosine methylation marks. In order to smooth analysis of EpiOne softwaer we recommend the use of a computer with adequate main and secondary memory.

Discussion

In order to validate EpiOne software in determining the distribution and frequency of potential DNA epigenetic marks, various promoters and gene bodies among different organisms consisting of mammals and plants were studied. Also whole mitochondrial and chloroplast genomes were also studied. Completion of the analyses was depended on the type of the analyses and the amount of data. Based on the χ^2 analysis [31] densities of potential DNA cytosine methylation marks CG, CHG and CHH were statistically different between promoters and gene bodies regardless of the organisms. The frequencies of CG among three sequence contexts (CpG, CpHpG, and CpHpH) were not well correlated with the length of genomic DNA sequence. On the other hand, there were moderate level simple positive correlation between the number of CpHpH and the genomic DNA lengths.

Analyses indicated that frequencies of three sequence contexts (CpG, CpHpG, and CpHpH) were statistically different within the gene bodies. For instance introns and 3'-UTR sequences contained similar CpHpH densities while densities of CpG statistically differed. These findings were supported with previous results in which hypomethylation was often associated with gene activation and hypermethylation was associated with gene repression [9]. The relationship between transcription level and gene body methylation level is currently not well understood. Therefore the use of EpiOne could provide some data to study the effect of gene body DNA methylation on transcriptional regulation.

Bisulfite sequencing is the gold traditional method for the detection of DNA methylation. High-throughput techniques still relies on traditional bisulfite sequencing, therefore, tools like EpiOne are still essential for the study of DNA cytosine methylation. In bisulfite sequencing methods and many other methods such as Southern blot detection of genomic regions from DNA material digested with restriction enzymes, methylation-specific PCR (MSP), methyl light or quantitative analysis of methylated alleles, methylation-sensitive single-stranded conformational polymorphism and bisulfite restriction analysis, require primers and amplification of the target regions for successful analysis [16,19,32,33]. Targets containing CpG, CpHpG or CpHpH and primer pairs flanking the targets could be selected using EpiOne software.

CG islands (CGIs) are stretches of DNA in which there is a higher than normal concentration of guanine and cytosine. Up to date multiple algorithms have been developed for identification of CG islands (CGIs). Detection of CGIs in genomic DNA sequences could be categorized into two groups: one is called the traditional algorithms based on three sequence parameters (length, GC content, and ratio of the observed over the expected CpG and the other algorithms based on statistical property in a sequence without imposing the three criteria in the traditional algorithms [34]. The EpiOne's algorithms could be used to determine CGIs in gene bodies and the regulatory sequences. This will further help in gathering the information about the distribution and the function of CGIs in organisms. Identification of CGIs is an initial step in many studies for identification of methylation status as well as in the design of methylation identification methods [9,34].

Conclusion

In conclusion algorithms embedded in EpiOne software could be used to study DNA cytosine methylation distribution among promoters and gene bodies and mine various genomic features, such as gene promoters, upstream and downstream regions (UTRs), transcription start site (TSS), coding exons or introns, and transcription termination sites (TTSs) for characterization of the genomic region. EpiOne software would be useful to investigate relationships between some gene properties such as gene length, exon, intron lengths and numbers. Promoter and gene body sequences extracted with EpiOne software can be used in development of molecular markers such as microsatellites, single nucleotide polymorphism, methylation-sensitive single-stranded conformational polymorphism and sequence tagged sites. We anticipate the use of EpiOne software will assist researchers to explore biological roles of DNA cytosine methylation in the epigenetic studies and related fields.

Acknowledgement

This software, EpiOne 1.0, was created as part of a project entitled "A research on CG, CHG, CHH methylation levels and epigenetic consequences in some plant groups: *in silico* and *in vitro* studies" funded by the Scientific and Technological Research Concil of Turkey (113O935). We acknowledge Assoc. Prof. Dr. M. Bilgen's help in the software programming.

References

[1]. Slack, J. M. (2002). Conrad Hal Waddington: the last Renaissance biologist? *Nature Reviews Genetics*, 3: 889-895.
[2]. Bender, J. (2004). DNA methylation and epigenetics. *Annual Review of Plant Physiology*, 55: 41-68.
[3]. Zemach, A., McDaniel, I., Silva, P., & Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*, 328: 916-919.
[4]. Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, 13: 484-492.
[5]. Plongthongkum, N., Diep, D. H., Zhang, K. (2014). Advances in the profiling of DNA modifications: cytosine methylation and beyond. *Nature Reviews Genetics*, 15: 647-661.
[6]. Mendizabal, I., Keller, T. E., Zeng, J., Yi, S. V. (2014). Epigenetics and evolution. *Integrative and Comparative Biology*, 54: 31-42.
[7]. Irizarry, R. A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., Ji, H., Potash, J. B., Sabunciyan, S., & Feinberg, A. P. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genetics*, 41: 178-186.
[8]. Hansen, K. D., Timp, W., Bravo, H. C., Sabunciyan, S., Langmead, B., McDonald, O. G., Wen, B., Wu, H., Liu, Y., Diep, D., Briem, E., Zhang, K., Irizarry, R. A., & Feinberg, A. P. (2011). Increased methylation variation in epigenetic domains across cancer types. *Nature Genetics*, 43: 768-775.
[9]. Lisanti, S., Omar, W. A. W., Tomaszewski, B., De Prins, S., Jacobs, G., Koppen, G., Mathers, J. C., & Langie, S. A. S. (2013). Comparison of methods for quantification of global DNA methylation in human cells and tissues. *PLoS ONE*, 8: e79044.
[10]. Sonkin, D., Hassan, M., Murphy, D., & Tatarinova, T. (2013). Tumor suppressors status in cancer cell line encyclopedia. *Molecular Oncology*, 7: 791-798.
[11]. Schroeder, D. I., Blair, J. D., Lott, P., Yu, H. O. K., Hong, D., Crary, F., Ashwood, P., Walker, C., Korf, I., Robinson, W. P., & LaSalle, J. M. (2013). The human placenta methylome. *Proceedings of the National Academy of Sciences* USA, 110: 6037-6042.
[12]. Finnegan, E. J., Peacock, W. J., & Dennis, E. S. (1996). Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development. *Proceedings of the National Academy of Sciences* USA, 93: 8449-8454.

[13]. Soppe, W. J., Jacobsen, S. E., Alonso-Blanco, C., Jackson, J. P., Kakutani, T., Koornneef, M., & Peeters, A. J. (2000). The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Molecular Cell*, 6: 791-802.
[14]. Chinnusamy, V., & Zhu J. K. (2009). Epigenetic regulation of stress responses in plants. *Current Opinion in Plant Biology*, 12: 133-139.
[15]. Groszmann, M., Greaves, I. K., Fujimoto, R., Peacock, W. J., & Dennis, E. S. (2013). The role of epigenetics in hybrid vigour. *Trends Genetics*, 29: 684-690.
[16]. Dhingra, T., Mittal, K., & Sarma, G. S. (2014). Analytical techniques for DNA methylation – an overview. *Current Pharmaceutical Analysis*, 10: 71-85.
[17]. Liu, J., & Jia, G. (2014). Methylation modifications in eukaryotic messenger RNA. *Journal of Genetics and Genomics*, 41: 21-33.
[18]. Goll, M. G., & Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annual Review of Biochemistry*, 74: 481-514.
[19]. Guerrero-Bosagna, C. (2013). DNA methylation research methods. *Materials and Methods*, 3: 206.
[20]. Ramsahoye, B. (2002). Measurement of genome wide DNA methylation by reversed-phase high-performance liquid chromatography. *Methods*, 27: 156-61.
[21]. Friso, S., Choi, S., Dolnikowski, G., & Selhub, J. (2002). A method to assess genomic DNA methylation using high-performance liquid chromatography/electrospray ionization mass spectrometry. *Analytical Chemistry*, 74: 4526-4533.
[22]. Kim, H., Park, J., Jeong, K., & Lee, S. (2007). Determining the global DNA methylation status of rat according to the identifier repetitive elements. *Electrophoresis*, 28: 3854-3861.
[23]. Pomraning, K. R., Smith, K. M., & Freitag, M. (2009). Genome-wide high throughput analysis of DNA methylation in eukaryotes. *Methods*, 47: 142-150.
[24]. Laird, P. W. (2010). Principles and challenges of genome wide DNA methylation analysis. *Nature Reviews Genetics*, 11: 191-203.
[25]. Yi, S. V., & Goodisman, M. A. (2009). Computational approaches for understanding the evolution of DNA methylation in animals. *Epigenetics*, 4: 551-556.
[26]. Krueger, F., & Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, 27: 1571-1572.
[27]. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Rapp, B. A., & Wheeler, D. L. (1997). GenBank. *Nucleic Acids Research*, 25: 1-6.
[28]. Karaca, M., Bilgen, M., Onus, A. N., Ince, A. G., & Elmasulu, S. Y. (2005). Exact Tandem Repeats Analyzer (e-TRA) for DNA sequence mining. *Journal of Genetics*, 84: 49-54.
[29]. Ince, A. G., Karaca, M., Bilgen, M., & Onus, A. N. (2008). Digital differential display tools for mining microsatellite containing organism, organ and tissue. *Plant Cell, Tissue and Organ Culture*, 94: 281-290.
[30]. Huang, D. W., Sherman, B. T, & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protocols*, 4: 44-57.
[31]. Ince, A. G. (2012). A contig-based microsatellite marker approach and its application in *Cichorium* ESTs. *Romanian Biotechnological Letters*, 17: 7177-7186.
[32]. Li, L. C., & Dahiya, R. (2002). MethPrimer: designing primers for methylation PCRs. *Bioinformatics*, 18: 1427-1431.
[33]. Aranyi, T., Varadi, A., Simon, I., & Tusnady, G. E. (2006). The BiSearch web server. *BMC Bioinformatics*, 7: 431.
[34]. Zhao, Z., & Han, L. (2009). CpG islands: algorithms and applications in methylation studies. *Biochemical and Biophysical Research Communications*, 382: 643-645.